

# The Lowest-Possible BER and FER for any Discrete Memoryless Channel with Given Capacity

Johannes B. Huber and Thorsten Hehn  
Institute for Information Transmission  
University of Erlangen-Nuremberg, Germany

## Abstract

We investigate properties of a channel coding scheme leading to the minimum-possible frame error ratio when transmitting over a memoryless channel with rate  $R > C$ . The results are compared to the well-known properties of a channel coding scheme leading to minimum bit error ratio. It is concluded that these two optimization requests are contradicting. A valuable application of the derived results is presented.

## I. INTRODUCTION

We consider coded data transmission over a memoryless channel with given capacity  $C$  for the case that the rate of the channel exceeds the channel capacity. This is a typical situation for a component code in a concatenated coding scheme [1]. The properties of a channel coding scheme with minimum average bit error ratio (BER) have been discussed in several papers (e.g. [2] and references therein). In this paper, we focus on schemes with minimum average frame error ratio (FER). The work presented in this paper is threefold. As a first contribution, we discuss a coding scheme optimal w.r.t. FER and use rate-distortion theory to derive the properties of the end-to-end channel. Second, we present a possible application of these findings. This is a lower bound on the rate when both the channel capacity and the tolerated average frame error ratio are specified. This leads to the most important and third contribution, the insight that minimum BER and minimum FER are contradicting targets which cannot be obtained by a single channel coding scheme. We show the consequences for the BER when the channel coding scheme is optimized w.r.t. to the FER and vice versa. The paper is organized as follows. Section II provides necessary definitions and describes the transmission system. In Section III we repeat the converse to the channel coding theorem which identifies a lower bound for reliable transmission. Section IV briefly repeats the properties of channel coding scheme with minimum BER, which was derived in [2]. In Section V ideal channel coding w.r.t. FER is introduced. An application as well as performance results for the non-optimized error ratio are shown in Section VI.

## II. TRANSMISSION SETUP AND DEFINITIONS

An information source delivers source symbols  $u[\ell]$ ,  $\ell = 1, \dots, k$ , from a binary alphabet. These symbols are realizations of the random variables  $U[\ell]$  which are assumed to be independent of each other and have identical distributions. In other words,  $H(U[\ell]) = 1$  holds, where  $H(\cdot)$  denotes the entropy of a random variable. In the following, we denote the  $j$ -th element of the set of possible source words by  $\mathbf{u}^{(j)}$ ,  $j = 1, \dots, 2^k$  and  $u[\ell]$ ,  $\ell = 1, \dots, k$  denotes the  $\ell$ -th entry in a vector  $\mathbf{u}$  of length  $k$ . An encoder of rate  $R = k/n$  is used to transform binary source vectors  $\mathbf{u}$  of length  $k$  into vectors  $\mathbf{x}$ . These vectors contain channel symbols and are of length  $n$ . The vectors  $\mathbf{x}$  are transmitted over the channel and received as vectors  $\mathbf{y}$  of length  $n$ , cf. Figure 1. Note that we do not assume any special properties of the channel except for being discrete and memoryless (discrete memoryless channel, DMC) and meeting the capacity  $C$ . A corresponding channel decoder uses the received vector  $\mathbf{y}$  to generate soft-output estimates of  $\mathbf{u}$ , denoted by  $\mathbf{v}$ . Binary quantization of  $\mathbf{v}$  yields  $\hat{\mathbf{u}}$ . As stated above, we are interested in channel coding schemes that minimize the BER and FER, respectively, when measured over the end-to-end channel. Here, the end-to-end channel corresponds to the channel transmitting  $\mathbf{u}$  to  $\mathbf{v}$  if soft-decision output is required and  $\hat{\mathbf{u}}$  otherwise, cf. Figure 1.

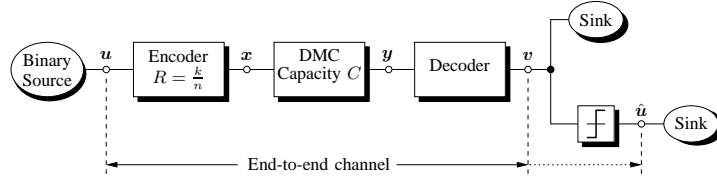


Fig. 1. Transmission scenario for signaling over a DMC of capacity  $C$

We define the average bit error ratio for a given position  $\ell$  as  $\text{BER}_\ell = \Pr(\hat{U}[\ell] \neq U[\ell])$ ,  $\ell = 1, \dots, k$  and the average bit error ratio in a codeword as  $\text{BER} = \frac{1}{k} \sum_{\ell=1}^k \text{BER}_\ell = \mathbb{E}_\ell\{\text{BER}_\ell\}$ . Additionally, a *tolerated* average bit error ratio  $\text{BER}_T$  is introduced. This ratio is technically equal to  $\text{BER}$ , as it is also measured between  $\mathbf{u}$  and  $\hat{\mathbf{u}}$ , but  $\text{BER}_T$  is a user-defined threshold variable. Henceforth, we tacitly assume  $\text{BER}_T \leq 0.5$ . Similarly, we define the average frame error ratio as the probability that the received frame differs from the transmitted one, i.e.  $\text{FER} = \Pr(\mathbf{U} \neq \hat{\mathbf{U}})$  where equality of two vectors is given if all elements in the two vectors are equal. Alike  $\text{BER}_T$ , we consider the tolerated average frame error ratio  $\text{FER}_T$ .

$I(\mathbf{U}; \mathbf{V})$  denotes the total mutual information between vectors (frames)  $\mathbf{u}$  and  $\mathbf{v}$  while  $I(U[\ell]; V[\ell])$  is the mutual information between an individual pair of input and output symbols. Additionally, we define

the average mutual information transmitted in a frame of  $k$  symbols,  $\bar{I}(U; V) = \frac{1}{k} \sum_{\ell=1}^k I(U[\ell]; V[\ell])$ .

### III. CONVERSE TO THE CHANNEL CODING THEOREM

We repeat the converse to the channel coding theorem as stated in [3, Ch. 4]. This theorem marks the starting point for both our considerations on the lowest BER and FER. In [3, Ch. 4] the transmission of a sequence of source digits is discussed. Note that these digits can represent both single information symbols as well as complete source words. We represent this distinction by different alphabets and as a consequence we denote the length of a sequence of source symbols by  $L$ . As each digit can be taken from an arbitrary alphabet, we denote the sequence of source digits by a sequence of vectors,  $\mathbf{u}_1^L = [\mathbf{u}_1, \dots, \mathbf{u}_L]$ . The sequence of channel digits and received digits, both of length  $N$ , are denoted by  $\mathbf{x}_1^N = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  and  $\mathbf{y}_1^N = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ , respectively. Please note that the cardinalities of the sets of channel input and output variables do not have to be specified, the mutual information  $I(\mathbf{X}_1^N; \mathbf{Y}_1^N)$  is sufficient. In the following, we consider the general error event that the source digit and the estimated digit do not coincide and denote the probability by  $P_e$ . Identifying the source digits from alphabets of size  $M = 2$  and  $M = 2^k$  allows us to deduct information on the average error ratios of interest, i.e. the BER and the FER, respectively.

We start by Equation (4.3.20) from [3] (Fano's inequality), which is a lower bound on the error probability  $P_e$  and reads

$$e_M(P_e) \geq \frac{1}{L} H(\mathbf{U}_1^L) - \frac{1}{L} I(\mathbf{X}_1^N; \mathbf{Y}_1^N). \quad (1)$$

Here,  $e_M(\cdot)$  is the  $M$ -ary entropy function,  $e_M(p) = e_2(p) + p \log_2(M-1)$ , and  $e_2(\cdot)$  is the usual binary entropy function  $e_2(p) = -p \log_2(p) - (1-p) \log_2(1-p)$ . Further,  $M$  denotes the size of the symbol alphabet.

Let us first assume that  $\mathbf{u}_1^L$  is a vector of binary symbols of length  $L$ , i.e.  $L = k$  and  $N = n$ . In this case,  $P_e$  coincides with the BER and  $\frac{1}{k} H(\mathbf{U}_1^k) = 1$  holds. Together with  $I(\mathbf{X}_1^n; \mathbf{Y}_1^n) \leq nC$  due to the memoryless channel and the capacity maximum of mutual information, the well known lower bound on BER results:

$$e_2(\text{BER}) \geq 1 - \frac{n}{k} C = 1 - \frac{C}{R}. \quad (2)$$

As mentioned above, we regard the source sequence  $\mathbf{u}$  for a lower bound on FER as one symbol out of a  $2^k$ -ary alphabet, and the variables  $\mathbf{x}$  and  $\mathbf{y}$  represent an entire codeword and received word, respectively.

In block coding, each source sequence is encoded into one codeword and subsequent codewords are mutually independent. Therefore,  $L = N$  holds and Equation (1) reads

$$e_{2^k}(\text{FER}) \geq \frac{1}{L}H(\mathbf{U}_1^L) - \frac{1}{L}I(\mathbf{X}_1^L; \mathbf{Y}_1^L) = H(\mathbf{U}) - I(\mathbf{X}; \mathbf{Y}).$$

With  $H(\mathbf{U}) = k$  and  $I(\mathbf{X}; \mathbf{Y}) \leq nC$ , we obtain the corresponding result to Equation (2) for FER,

$$e_{2^k}(\text{FER}) \geq k - nC = k(1 - \frac{C}{R}). \quad (3)$$

It is worth mentioning that Equation (2) is a special case of Equation (3) for  $k = 1$ . In the following, we will show by means of rate-distortion theory that the lower bounds in Equation (2) and Equation (3) can be met with equality by optimized channel coding schemes. We will denote the average frame error ratio measured in a system optimized w.r.t. to minimum BER by  $\text{FER}'$ . Likewise, the average bit error ratio measured in a system optimized w.r.t. to minimum FER will be denoted by  $\text{BER}'$ .

#### IV. OBTAINING THE LOWEST POSSIBLE BER FOR A MEMORYLESS CHANNEL WITH GIVEN CAPACITY

We investigate the transmission of data at a rate which exceeds the capacity, i.e. we consider the region where error-free transmission is not possible. Rate-distortion theory [4] postulates, that if an end-to-end average bit error ratio  $\text{BER}_T$  is tolerated, a code with rate  $R$  and appropriate decoding rule exists and achieves an average bit error ratio  $\text{BER} \leq \text{BER}_T$  as long as

$$R \leq \frac{C}{1 - e_2(\text{BER}_T)} \quad (4)$$

and if  $n \rightarrow \infty$ .

We define a coding scheme (i.e. code, encoder, and decoder) with rate  $R = \frac{C}{1 - e_2(\text{BER}_T)}$  to be *ideal in terms of the average bit error ratio*, iff the average bit error ratio meets the tolerated one,  $\text{BER} = \text{BER}_T = e_2^{-1}(1 - \frac{C}{R})$ . It was shown in [2] that Equation (4) is met with equality when signaling over a memoryless binary symmetric channel (BSC).

This leads to the conclusion that the use of a coding scheme ideal w.r.t. BER results in an end-to-end channel which is a *memoryless* BSC [2]. For completeness we add that in [2] it is also shown that for an ideal coding scheme w.r.t. BER,  $\bar{I}(U; \hat{U}) \equiv \bar{I}(U; V) \equiv \frac{1}{k}I(\mathbf{U}; \mathbf{V})$  holds, i.e. soft-output has no benefit over hard-output and interleaving has no influence on the memoryless sequence of errors.

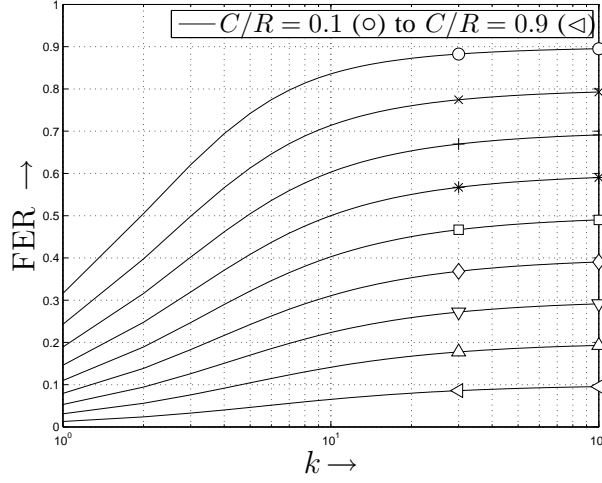


Fig. 2. Minimum FER for different block lengths and given  $C/R$

## V. OBTAINING THE LOWEST POSSIBLE FER FOR A MEMORYLESS CHANNEL WITH GIVEN CAPACITY

In this section we present the first contribution of this paper. We discuss a channel coding system leading to the end-to-end channel with the lowest possible average frame error ratio. The adaptation of the converse to the channel coding theorem for the FER being the error ratio is given by Equation (3).

Figure 2 shows the corresponding lower bound on the FER over  $k$  for different values of  $C/R$ . The lower bound on  $C/R$ , specified by Equation (3), becomes particularly interesting when  $k$  approaches large values. Then,  $C/R \geq \lim_{k \rightarrow \infty} \left( 1 - \frac{e_2(\text{FER})}{k} - \frac{\log_2(2^k - 1)}{k} \text{FER} \right) = 1 - \text{FER}$ , or equivalently,

$$\text{FER} \geq 1 - C/R. \quad (5)$$

Again, we show that this bound can be met with equality by means of rate-distortion theory and by this it is proven that a coding scheme meeting the FER has indeed to exist [5]. To this end, let us first show that there exists a channel meeting inequality (5) with equality, namely the  $M$ -ary symmetric channel ( $M$ -SC) with  $M = 2^k$ . We assume that  $\mathbf{u}$  and  $\hat{\mathbf{u}}$  denote the input and output symbols of that channel, respectively. The transition probabilities of this end-to-end channel are denoted as  $\Pr(\mathbf{U} = \mathbf{u}^{(j)} | \hat{\mathbf{U}} = \mathbf{u}^{(j)}) = 1 - \text{FER}$   $\forall j \in \{1, \dots, 2^k\}$  and  $\Pr(\mathbf{U} = \mathbf{u}^{(i)} | \hat{\mathbf{U}} = \mathbf{u}^{(j)}) = \text{FER}/(2^k - 1)$ ,  $\forall j, j \neq i$ .

The mutual information per channel use is calculated by

$$I(\mathbf{U}; \hat{\mathbf{U}}) = H(\mathbf{U}) - H(\mathbf{U} | \hat{\mathbf{U}}) \Big|_{\Pr(\mathbf{U})=[2^{-k} \dots 2^{-k}]} = \quad (6)$$

$$k + \sum_{i=1}^{2^k} \sum_{j=1}^{2^k} \Pr(\hat{U} = \mathbf{u}^{(j)} | U = \mathbf{u}^{(i)}) \Pr(U = \mathbf{u}^{(i)}) \log_2 \left( \frac{\Pr(\hat{U} = \mathbf{u}^{(j)} | U = \mathbf{u}^{(i)}) \Pr(U = \mathbf{u}^{(i)})}{\sum_{i'} \Pr(\hat{U} = \mathbf{u}^{(j)} | U = \mathbf{u}^{(i')}) \Pr(U = \mathbf{u}^{(i')})} \right),$$

what can be simplified to

$$I(U; \hat{U}) = k - e_{2^k}(\text{FER}). \quad (7)$$

Equation (7) denotes the mutual information for the transmission of a whole vector. Normalized to one binary symbol it reads  $\frac{I(U; \hat{U})}{k} = 1 + \frac{(1-\text{FER})}{k} \log_2(1 - \text{FER}) + \frac{\text{FER}}{k} \log_2 \left( \frac{\text{FER}}{2^k - 1} \right)$ , and thus for  $k \rightarrow \infty$ :

$$\frac{I(U; \hat{U})}{k} = 1 - \text{FER}.$$

Considering the data processing theorem in the form  $\frac{I(U; \hat{U})}{k} \leq \frac{C}{R}$  and Equation (5) in the form  $C/R \geq 1 - \text{FER}$  allows us to conduct  $\frac{I(U; \hat{U})}{k} = \frac{C}{R} = 1 - \text{FER}$  for  $k \rightarrow \infty$ . Again, by making use of rate-distortion theory and considering the fact that a distinct test channel exists [5], we conclude that the lower bound provided in Equation (5) can be met with equality.

In the following, we will review the properties of this channel. In the case of an error, the  $2^k$ -ary symmetric channel maps the input to all incorrect outputs with equal probability. We therefore conclude that if a frame error occurs, the average bit error ratio within these frames is 0.5. Hence, the resulting end-to-end channel corresponds to a fully bursty channel. More strictly speaking, a block-erasure channel with average erasure probability FER and infinite frame length meets the bound  $C/R \geq 1 - \text{FER}$  with equality. This finding allows us to establish a coherence between the capacity and the rate, when an average frame error ratio  $\text{FER}_T$  is tolerated. This coherence reads  $R \leq \frac{C}{1 - \text{FER}_T}$ .

We define a coding scheme (i.e. code, encoder, and decoder) with rate  $R = \frac{C}{1 - \text{FER}_T}$  to be *ideal in terms of the average frame error ratio*, iff the average frame error ratio meets the tolerated frame error ratio  $\text{FER}_T$  with equality,  $\text{FER} = \text{FER}_T$ .

We denote the obtained average bit error ratio of such a channel by  $\text{BER}'$ . There exists a straightforward coherence between  $\text{BER}'$  and the optimal average frame error ratio FER which reads  $\text{BER}' = \frac{1}{2} \left( 1 - \frac{C}{R} \right)$ . The capacity of the fully bursty binary (end-to-end) channel, where all errors are part of very long error bursts, can be written as  $C = 1 - 2\text{BER}'$ . This is due to the fact that all errors are concentrated in bursts and within these bursts the average bit error ratio is 0.5. Reliable communication is accomplished by the simple rule of erasing the error bursts at the receiver side. For error detection, e.g. by means of a cyclic

redundancy check (CRC), additional redundancy is necessary but this cost vanishes for  $k \rightarrow \infty$ . Alike stated in Section IV, we observe that soft information has no benefit over hard output if an optimal coding scheme w.r.t. minimum FER is used. Examples for such fully bursty channels can simply be generated by renewal burst channel models, like the model of Fritchman with a single error state [6]. For given BER, the capacity of such a channel is maximized when the average burst length tends to infinity and in this limit, the capacity equals  $1 - 2\text{BER}$ . This entity exactly corresponds to the situation of bit errors at the output of a coding scheme which is ideal w.r.t. minimum average frame error ratio. Consider now an end-to-end channel with minimum BER. For  $k \rightarrow \infty$ , the obtained average frame error ratio, denoted by  $\text{FER}'$  is given by  $\text{FER}' = 0$  iff  $\text{BER} = 0$  and  $\text{FER}' = 1$  otherwise. When considering blocks of infinite length, every block is erroneous if bit errors are possible in general.

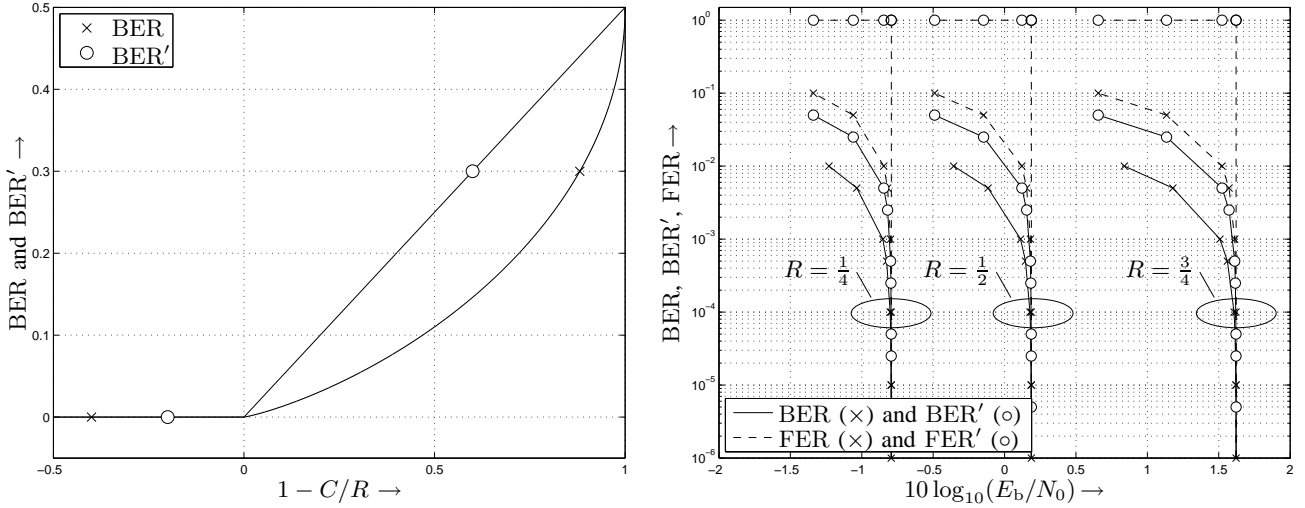
With the results derived so far, it is straightforward to see that a channel coding scheme working in the region  $R > C$  cannot obtain the minimum-possible BER and the minimum possible FER with one channel coding scheme, cf. Figure 3(a).

## VI. POSSIBLE APPLICATION

A possible application of the presented results is introduced in this section. We assume binary antipodal signaling (BPSK) over the AWGN channel with a channel code of given rate  $R$ . A lower bound on the obtainable BER is given in Equation (2), which can be rewritten to  $\text{BER} \geq e_2^{-1} \left(1 - \frac{C}{R}\right)$ .

This entity allows to generate the curves depicting the optimum BER obtainable by codes of given rate and length approaching infinity. These curves are well-known from numerous publications within the area of channel coding and visualize the fundamental limits for transmission at a given rate. In Figure 3(b) these curves are shown for the rates  $R = 1/4$ ,  $R = 1/2$ ,  $R = 3/4$ , respectively, and are labeled by BER. Here, the capacity of the channel is specified by the signal-to-noise ratio expressed by  $10 \log_{10}(E_b/N_0)$  as usual. In this context,  $E_b$  denotes the energy per transmitted bit of information and  $N_0$  represents the one-sided spectral noise-power density.

The findings presented in Section V allow to extend these fundamental curves to scenarios where the FER is used as the performance measure. Here, Equation (5) states the lower bound on the FER which can be reached by code lengths approaching infinity. Figure 3(b) also depicts lower bounds on the frame error ratios for codes of rate  $R = 1/4$ ,  $R = 1/2$ , and  $R = 3/4$ . These curves are labeled by FER. Figure 3(b) also shows the average error ratios  $\text{BER}'$  and  $\text{FER}'$ . These curves illustrate the average bit error ratio for the case that the channel coding system of the end-to-end channel has been optimized for



(a) Visualization of the significant difference between BER and BER' for coding schemes optimal w.r.t BER and FER, respectively. (b) BER, BER', FER and FER' for coding schemes of given rate.

Fig. 3. Performance comparisons on BER, BER', FER and FER'

the FER and the average frame error ratio for the case that the optimization was done for the BER, respectively. This confirms that channel coding schemes being ideal w.r.t. minimum BER and minimum FER have to be designed in quite different ways. Especially an optimization w.r.t. the BER in situations, when a low average frame error ratio is required as well, leads to significant performance losses. In all cases, the average mutual information of the end-to-end channel is given by  $\min(C/R, 1)$ . In order to obtain minimum BER, the end-to-end channel is a memoryless BSC, whereas for minimum FER, an end-to-end channel with memory (to be precise, a block-erasure channel) results.

## VII. CONCLUSIONS

We considered transmission at rates exceeding the capacity of the underlying channel. Fundamental insight in a threefold manner is given. First, knowledge on a coding scheme leading to an end-to-end channel with minimum average frame error ratio is provided. It turns out that this channel is a block-erasure channel, transmitting frames either correctly or in such a way that no information is transmitted at all. The average bit error ratio within a frame corresponding to a burst error equals 0.5. Second, it was shown that minimum BER and minimum FER are disparate requests to a channel coding scheme. The third contribution is an application. It is usual in literature to compare the BER behavior of channel coding schemes to information theoretic bounds; this is now also possible with respect to the FER.



## ACKNOWLEDGMENT

*The authors want to thank the anonymous reviewers for their valuable comments.*

## REFERENCES

- [1] S. ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications*, 49(10):1727–1737, October 2001.
- [2] S. Huettinger, J.B. Huber, R.F.H. Fischer, and R. Johannesson. Soft-output-decoding: Some aspects from information theory. In *Proceedings of the International ITG Conference on Source and Channel Coding (SCC)*, pages 81–89, Berlin, Germany, January 2002.
- [3] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, 1968.
- [4] C.E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record*, 4:142–163, 1959.
- [5] Toby Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, N.J., January 1971.
- [6] B. Fritchman. A binary channel characterization using partitioned Markov chains. *IEEE Transactions on Information Theory*, 13(2):221–227, April 1967.